Frame Semantics for EU Reporting Requirements: An Annotated Benchmark, Guidelines, and a Path Towards RRMV Population

Célestine Juliette Marie Lecat[1], Kimmy Shah[2] and Monica Palmirani[3]

[1]CIRSFID ALMA-AI, Department of Legal Studies University of Bologna, Bologna, Italy
[2]LSTS Research Group, Vrije Universiteit Brussel, Brussels, Belgium
[3]CIRSFID ALMA-AI, Department of Legal Studies, University of Bologna, Bologna, Italy

Abstract

Reporting Requirements are particular legislative provisions that require certain institutions (e.g., the EU Commission, Member States, agencies) to prepare a report, plan, budget, opinion, or standard to address other institutions for monitoring the quality of legislation. Identifying and monitoring these particular types of obligations is currently a largely manual, time-consuming task, and EU regulatory-simplification efforts call for a reduction of the red-tape burden. We present a frame-semantic annotation layer over a corpus of reporting requests, building on the corpus introduced by Corazza et al. [1] within the SORTIS (Study on Regulatory Reporting Standards) project. Using FrameNet, we annotate each reporting request with six frames: `Telling`, `Intentionally_create`, `Conditional_scenario`, `Time_vector`, `Frequency`, and `Artifact`. These frames capture who must report what, to whom, under which conditions, by when, and how often. We describe our annotation guidelines and an inter-annotator agreement study on a second annotator with no FrameNet knowledge, finding perfect frame- and role-label agreement whenever both annotators mark the same span ($\kappa = 1.000$ and $\kappa = 0.994$ respectively) and identifying two concrete, fixable guideline gaps behind the remaining disagreement. We propose a mapping from our frame elements onto the Reporting Requirement Metadata Vocabulary (RRMV) ontology, illustrated with an example. The result is a small reusable resource intended to support future ML/LLM-based extraction of reporting requirements and their semi-automatic population into machine-readable, queryable legal knowledge graphs. The goal is to improve the detection of semantic elements at a more granular level, particularly in sentences with unusual wording that deviates from more common linguistic patterns (e.g., temporal conditions).

Keywords

Frame Semantics, FrameNet, Legal Informatics, Reporting Requirements, RRMV, Semantic Annotation

CEUR-WS.org/Vol-4327/paper4.pdf

1. Introduction

EU legislation frequently contains provisions that require an institution or Member State (the *Addresser*) to supply a document to another EU body (the *Addressee*), typically the European Commission, the Council, or the European Parliament. We refer to such provisions as *reporting requests* (RRs). RRs are found across EU legislation (e.g., Directives, Regulations, implementing acts), and their number and complexity have grown substantially in recent years, increasing the burden on the EU institutions that must track overlapping mandates[1]. On top of this, the underlying legislation itself keeps evolving through amendments and derogations over time [2]. As a consequence, tracking RR mandates has become a difficult task that is still performed largely through manual legal review. This tension is not only academic: it aligns with the EU's ongoing regulatory simplification agenda, which presupposes that RRs and their components can first be identified at scale.

This paper takes a small step towards that goal. We do not attempt to solve automatic RR extraction. Instead, we build the kind of linguistically grounded, human-annotated resource that extraction systems, whether classical ML or LLM-based, would need for training and evaluation.

It is reasonable to ask why a frame-semantic layer is needed at all, when an LLM could be prompted to produce RRMV instances from legislative text directly. We see three arguments for the intermediate layer. First, *independence from the target ontology*: the annotation records linguistic structure rather

NXDG 2026: 3rd NeXt-Generation Data Governance Workshop, colocated with SEMANTiCS, September 15-17, 2026, Ghent, Belgium
celestine.lecat2@unibo.it (C. J. M. Lecat); kimmy.shah@vub.be (K. Shah); monica.palmirani@unibo.it (M. Palmirani)
0009-0008-7109-7598 (C. J. M. Lecat); 0000-0002-8557-8084 (M. Palmirani)
© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[1]https://www.europarl.europa.eu/RegData/etudes/STUD/2024/762858/EPRS_STU(2024)762858_EN.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-11-13

than commitments to a particular vocabulary, so it survives revisions of the target ontology. A system producing RRMV directly must be re-run and re-validated against each new release, whereas the frame layer needs only to be re-mapped. Second, *reuse across languages*: frames and frame elements are defined independently of any particular language, which matters for a body of law that exists in 24 official languages. Third, *localisable evaluation*: frame-semantic parsing decomposes into three subtasks with existing parsers and benchmarks [3, 4], so an incorrect assertion can be attributed to a specific stage, rather than only observed as a wrong output. We stress that we have not shown the frame-mediated pipeline to outperform direct extraction, which we leave to future work. In either case, a corpus of the present kind is a prerequisite for evaluating such systems at all. What the present study does show is that an annotator with no prior exposure to FrameNet can reproduce this intermediate representation with complete frame-label agreement (Section 4). We see this as a first step: a small, carefully checked resource that can later be scaled up, incorporated into a broader semantic stack (Akoma Ntoso LegalDocML XML vocabulary [5], ELI (European Legislation Identifier [2]), and used to train and evaluate automatic extraction systems for reporting requirements. This work is guided by three research questions:

RQ1 Can frame semantics provide a fine-grained but reproducible representation of the linguistic structure of EU reporting requests? We contribute a frame-semantic annotation layer built with FrameNet [6] over the corpus of Corazza et al. [1], with detailed annotation guidelines.

RQ2 Is the annotation scheme reliable and reproducible by other annotators, given only the guidelines? We contribute an inter-annotator agreement study on 19 reporting requests, annotated blind by a second annotator with a legal background but no FrameNet knowledge (Section 4).

RQ3 Can this linguistic layer be mapped onto an existing legal knowledge representation for reporting requirements (RRMV), and what does this attempt reveal? We contribute an explicit mapping from our frame elements to RRMV classes and properties [7, 8], with a worked RDF example and an account of where this mapping is clean and where it exposes gaps.

2. Background

2.1. Reporting requirements, the SORTIS project and RRMV

The corpus underlying this work was introduced by Corazza et al. [1] as part of the SORTIS project, which focuses on systematically identifying and formalising reporting requirements within EU legislation using a hybrid AI approach combining rule-based and machine-learning/LLM-based extraction. Their work also introduced the Regulatory Reporting Metadata Vocabulary (RRMV) [8, 7] as the target representation for extracted reporting requests, enabling their tracking over time and structured querying. RRMV is an OWL[3] ontology for modelling RRs in legal resources, built on top of ELI (European Legislation Identifier) [9], FRBRoo[4], Dublin Core Terms[5], and the Web Annotation Data Model [10]. We work against **Release 1.0.0**, published as a SEMIC Recommendation on 2025-10-01 [7].

The present paper takes the SORTIS corpus as its starting point and adds a fine-grained, frame-semantic layer on top of it. It is worth noting that a reporting requirement (also called regulatory reporting[6]) is, strictly speaking, not a legal obligation in the technical sense, since failure to report does not by itself trigger a violation or a sanction [8]. Functionally, however, it is treated as binding throughout the EU's Better Regulation strategy for monitoring policy implementation and legislative simplification.

[2] https://eur-lex.europa.eu/content/help/eurlex-content/eli.html?locale=en

[3] OWL 2 Web Ontology Language Document Overview, W3C Recommendation: https://www.w3.org/TR/owl2-overview/

[4] FRBR-object oriented (FRBRoo), jointly maintained by the CIDOC CRM Special Interest Group and the IFLA FRBR Review Group: https://cidoc-crm.org/frbroo/home-0

[5] DCMI Metadata Terms, DCMI Recommendation: https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

[6] European Commission, Interoperable Europe portal, glossary entry "Regulatory reporting": https://interoperable-europe.ec.europa.eu/collection/better-legislation-smoother-implementation/glossary/term/regulatory-reporting

2.2. FrameNet in the legal domain

Frame semantics [11] model word meaning in terms of *semantic frames*: structured representations of a situation type combined with the relevant participant roles, called *frame elements* (FEs). The theory was operationalised at scale by the Berkeley FrameNet[7] project [6, 12], and its annotation methodology and conventions are documented in detail in the FrameNet II manual [13]. A word or expression that evokes (or "triggers") a frame is called a *lexical unit* (LU). FrameNet annotation is a three-step task: (i) identifying the LU that triggers a frame, (ii) identifying the text spans that fill that frame's elements, and (iii) labelling each span with the correct FE. We based our annotation of RRs on these three steps: each RR can be represented with frame-evoking predicates labeled as LUs (*develop*, *submit*, *report*, ...) together with the participants and circumstances (FEs) that FrameNet already has a vocabulary for. Only a subset of frames and FEs were selected for the annotation, with minor customisation.

Applying FrameNet to legal text is not new, and our work builds directly on this line of research rather than starting from scratch. Palmirani et al. [14] used a FrameNet approach to model the suspension of legal norms in Italian legislation, isolating temporal parameters through a combination of deep parsing and shallow semantics. Our use of the `Time_vector` frame is similar to this use case. Venturi [15] carried out a broader empirical trial annotating a corpus of Italian legal texts with FrameNet, arguing that the approach is well suited to making the linguistic and semantic content of legal provisions explicit and to supporting automatic semantic processing downstream. Agnoloni et al. [16] went a step further and directly compared a FrameNet-style description against an ontological characterisation of a Fundamental Legal Concept (*obligation*), investigating how far the two representations can be mapped onto one another. We drew inspiration from this work for the RRMV mapping we present in Section 5.

3. Corpus and Annotation Scheme

The corpus consists of reporting requests drawn from EU legislation, each provision containing at least one command to supply a report-like artifact to an EU institution. The corpus comprises 91 RRs annotated in English using the INCEpTION platform [17]. The guidelines were iteratively tailored to the corpus by the reference annotator. Three working days were necessary for the reference annotator to mark up the full corpus, and one working day was needed for the second annotator to work on the sample corpus (19 RR). Once the reference annotation was done, we ran a scripted corpus-wide consistency check using Anthropic's Claude (Pro) to spot invalid POS names, invalid role-for-frame combinations, missing core roles, orphan spans, discontinuous spans, Landmark/Distance misassignment, bare Frequency... and manually evaluated and fixed the issues that surfaced across all 91 provisions.

3.1. Three annotation layers

Following standard frame-semantic parsing practice, each reporting request is annotated on three coordinated layers in INCEpTION:

1. **Lexical Unit (LU)**: the word(s) evoking a frame, tagged with a `lemma.POS` value (e.g. `develop.v`, `by.prep`). Multiword expressions functioning as a single unit (e.g. `not_later_than.conj`) are tagged uniformly across their tokens.
2. **Frame**: co-extensive with the LU span, labelled with the frame name.
3. **Frame Element (FE)**: one span per argument, linked to a role from the evoking frame's role list.

3.2. The six frames

`Telling` The reporting act itself: a *Speaker* addresses an *Addressee* with a *Message* (optionally a *Topic* and/or *Medium*). We deliberately broaden this frame beyond canonical communication verbs (*submit*, *inform*, *disclose*, *publish*, *report*) to cover any verb describing a document being delivered or attached to a recipient (e.g. *accompany*), since such cases function identically for our purposes.

[7] https://framenet.icsi.berkeley.edu/intro

`Intentionally_create` The prior act of producing the reported artifact, when the text requires it to be developed/drafted, with *Creator*, *Created_entity*, and *Purpose*.

`Conditional_scenario` The condition(s) (*Profiled_possibility*) under which a *Consequence* (the reporting obligation) applies.

`Time_vector` The deadline or temporal constraint on the reporting *Event*, via a *Direction* (before/after) relative to a *Landmark_event* and/or an explicit *Distance*.

`Frequency` How often the reporting *Event* recurs.

`Artifact` The reported object itself and, where relevant, its part-whole structure (*Artifact_subpart*, *Artifact_superpart*). The FEs *Artifact_superpart* and *Artifact* have been added for more granularity.

3.3. Key annotation conventions

Several conventions keep the scheme internally consistent. FE spans are always contiguous but may be nested. The same FE span can be linked to roles from more than one frame when the same text genuinely fills both (e.g. a report noun phrase serving as both `Telling`'s *Message* and `Artifact`'s *Artifact*). Coordinated fillers ("the Parliament, the Council and the Commission") receive one FE span and one role link *per conjunct*, rather than a single span over the whole coordination. For `Time_vector`, a bare calendar date or named reference point with no explicit numeral duration is tagged *Landmark_event*; an explicit numeral+unit duration is tagged *Distance*. For `Frequency`, the *Event* FE is always linked to the governing predicate or event-denoting noun.

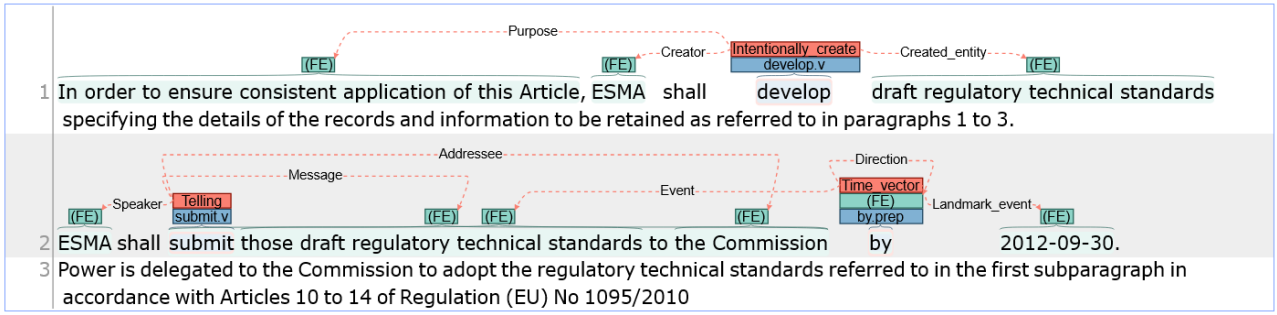


Figure 1: An example of a marked up reporting request (RR_0015).

4. Inter-Annotator Agreement

To assess the reproducibility of the annotation scheme, a second annotator independently annotated a sample of 19 reporting requests, randomly selected from the corpus. The reference annotator is a linguist accustomed to FrameNet. The second annotator has a background in law and technology and is a native speaker of English, but has no prior familiarity with FrameNet or frame-semantic annotation. The second annotator was given a locally-run INCEpTION project, configured with the same layers and tagsets as the reference project (Section 3), containing 24 reporting requests in total: 5 pre-annotated examples provided for reference alongside the guidelines, and the 19 reporting requests to be annotated. Annotation was carried out using only the guidelines and the worked examples, without any communication with the reference annotator.

4.1. Metrics

We report agreement separately for the three sub-tasks of frame-semantic parsing [3] – frame identification, argument identification, and role classification – to avoid conflating different sources of

disagreement. Agreement is quantified using Cohen's κ [18], following the standard methodology for inter-coder agreement in computational linguistics laid out by Artstein and Poesio [19], together with F_1 and Jaccard for set-based comparisons:

- **Frame identification**: Cohen's κ over frame labels at matched trigger positions, plus raw agreement on whether a frame is evoked at all at a given position;
- **Argument (span) identification**: pairwise F_1 between the two annotators' FE spans, computed both under exact-match and under a relaxed partial-overlap criterion, given the small sample size;
- **Role classification**: Cohen's κ over role labels, computed only on the subset of spans both annotators agree exist (so that span disagreement does not contaminate role disagreement).

We compute these metrics with a dedicated script defined with Anthropic's Claude (Pro) rather than INCEpTION's built-in agreement view. The latter offers per-layer Cohen's κ, Fleiss' κ, and Krippendorff's α out of the box, but by its own documentation does not support agreement calculation over stacked annotations.

4.2. Results

Table 1
Inter-annotator agreement on 19 shared reporting requests

Sub-task	Metric	Value
Target identification (raw trigger detection)	Jaccard index	0.759 (66/87)
Frame identification (in intersection of detected triggers)	Cohen's κ	1.000 (66/66)
Argument identification (exact match)	F_1	0.607
Argument identification (relaxed match, Dice ≥ 0.5)	F_1	0.824
Role classification (in intersection of detected FEs)	Cohen's κ	0.994 (173/174)

Target identification (0.759) is reported as raw set overlap (Jaccard) rather than a chance-corrected coefficient. Cohen's κ would require a well-defined negative class, i.e. a count of positions where both annotators agreed no frame was present. However, every token in the corpus is a potential trigger, so that denominator would be arbitrary (thousands of tokens, the agreement would inflate toward 1.000 no matter what).

Frame identification, by contrast, is computed over a restricted item set: the 66 positions *both* annotators marked. Here Cohen's κ is appropriate, and perfect frame-label agreement was achieved (66/66, $\kappa = 1.000$). This indicates that the six frames are cleanly separable: when both annotators looked at the same target, there was no ambiguity about which of the six it evoked. Role-label agreement on matched spans for FEs is also near perfect ($\kappa = 0.994$; the single exception is *Distance* vs. *Direction* on the immediacy adverb "immediately", which does not cleanly fit either FE's current definition).

4.2.1. Inter-annotator disagreement

Disagreement instead concentrates in span boundaries (exact-match $F_1 = 0.607$, rising to $F_1 = 0.824$ once near-miss boundaries are tolerated) and in whether a frame should be evoked at a given position at all (raw trigger-detection agreement 0.759). Regarding the latter, 4 of the 18 frame instances added only by the second annotator share a single construction: a verb (*specifying*, *setting out*, *assessing*) introducing an enumerated list of report contents, which is tagged `Artifact` by the second annotator but left unannotated in the reference annotation. The cause is a specific and fixable guideline gap.

Aside from this, seven of the eighteen frame additions stem from (1) an error from the reference annotator (RR_0084) and (2) a general, recurring disagreement: the nature of the provisions themselves, RR or not RR (e.g. RR_0011)? In both cases the disagreement reflects an (conscious or unintentional) omission in the reference annotation rather than an error by the second annotator, and in both cases a

single missed provision inflates the instance-level count several-fold. The uncertainty of the nature of provisions is an incentive to further define the notion of reporting request.

In parallel, only three frame instances appear in the reference annotation but not in the control annotation, each reflecting a different phenomenon. For example, in RR_0076, the reference annotator annotated two separate `Telling` frames for coordinated verbs ("**inform** and **forward**"), whereas the second annotator treats the coordination as a single communication event with a compound Message; and in RR_0038, the temporal indication "without delay" is annotated as a `Time_vector` in the reference annotation but not in the control. Segmentation of coordinated predicates, passive forms and immediacy verbs are points that should be further specified in the guidelines.

5. Towards Ontology Population: Mapping to RRMV

Beyond serving as training/evaluation data, this annotation layer could be useful for populating the RRMV ontology. RRMV already defines a `hasTarget` property linking an `oa:Annotation` to the model elements it defines. Our frame/FE spans, together with their exact character offsets, are a natural source of such annotations. Table 2 summarises the main correspondences we identified between our FEs and RRMV constructs (the complete mapping is provided as supplementary material).

Table 2
Selected Frame/FE to RRMV mappings (abridged; full table in supplementary material)

Frame element	RRMV construct	Fit
Telling.Addressee	AgentRole.hasType = `:addressee`	clean
Telling.Speaker	AgentRole.hasType = `:addresser`	clean
Intentionally_create (event) → Telling (event)	Action.hasNext	clean
Time_vector. Landmark_event	InstantSpec	clean
Frequency.Event	TemporalEntitySpec.hasFrequency → Frequency(value, hasUnit)	clean
Conditional_scenario. Profiled_possibility	—	gap
Artifact_subpart/superpart	—	gap

5.1. Clean fits, and gaps as a contribution

RRMV's role-type vocabulary maps cleanly onto our Addresser/Addressee distinction. RRMV's worked example (Member States reporting greenhouse gas inventories to the Commission under Article 26(2) of Regulation (EU) 2018/1999) is structurally close to our corpus and assigns `rrmv:addresser` to the reporting Member State, i.e., the same sense in which FrameNet's `Telling.Speaker` uses "Addresser."

The clearest validation of the mapping comes from the temporal module: `Time_vector` and `Frequency` map almost attribute-for-attribute onto RRMV's `TemporalEntitySpec` hierarchy (`InstantSpec`, `DurationSpec`, `before/after`, and `Frequency.value/hasUnit`), and the two-step "develop, then submit" pattern that recurs across the corpus maps cleanly onto RRMV's `Action.hasNext` chaining, both confirmed directly against Release 1.0.0's own worked examples rather than inferred. This exercise highlights gaps in the current RRMV vocabulary: no property for the communication *Medium* or for an artifact's *Purpose*, no `Condition` class or property capturing `Conditional_scenario`'s conditional applicability, no part-whole property scoped to `ActionResult`, and no `ResultType` instances for *opinion* or *proposal* artifacts present in our corpus, all absent in Release 1.0.0. The identification of these gaps, grounded in real annotated text, is a modest but concrete contribution.

5.2. Worked example

The abridged listing below shows a worked, hand-checked instantiation of RRMV 1.0.0 for a two-sentence reporting request in our corpus (RR_0002), including the linguistic grounding of each RRMV individual via `oa:Annotation` instances that carry the exact character offsets from the underlying WebAnno TSV export. The full listing is provided as supplementary material (`rr_0002_rrmv_example.ttl`)[8].

```
ex:RR_0002 a rrmv:Requirement ;
    rrmv:hasURI "http://example.org/corpus/RR_0002"^^xsd:anyURI ;
    rrmv:produces ex:Action_Create .

ex:Action_Create a rrmv:Action ;
    rrmv:hasAgentRole ex:AgentRole_Creator ;
    rrmv:hasResult ex:Result_Standard ;
    rrmv:hasNext ex:Action_Submit .

ex:Role_Addresser_Create a rrmv:Role ; rrmv:hasType rrmv:addresser .
ex:Result_Standard a rrmv:ActionResult ; rrmv:hasType rrmv:standard ;
    dcterms:title "draft regulatory technical standards"@en .

ex:Action_Submit a rrmv:Action ;
    rrmv:hasAgentRole ex:AgentRole_Speaker, ex:AgentRole_Addressee ;
    rrmv:hasResult ex:Result_Standard ;
    rrmv:atTime ex:Landmark_Date .

ex:Landmark_Date a rrmv:InstantSpec ; rrmv:hasFunction rrmv:deadline ;
    rdfs:label "2018-07-21"@en .

ex:Anno_Addressee a oa:Annotation ;
    rrmv:hasTarget ex:AgentRole_Addressee ;
    oa:bodyValue "the Commission"@en ;
     oa:hasSelector [
        a oa:XPathSelector ;
        rdf:value "//akn:role[@eId='role_12']"
    ]  .
```

6. Future Work and Conclusion

We presented a frame-semantic annotation layer over a corpus of EU RRs, using six FrameNet frames to make their content explicit. On **RQ1**, the six frames and their conventions show that FrameNet's existing inventory is expressive enough to capture Addresser, Addressee, content, condition, deadline, and frequency without a bespoke scheme. On **RQ2**, frame- and role-label agreement are near-perfect whenever both annotators identify the same span ($\kappa = 1.000$ and $\kappa = 0.994$ respectively), and several of the remaining disagreements trace to a few specific, fixable guideline gaps. On **RQ3**, several frames map cleanly onto RRMV Release 1.0.0 while others expose concrete gaps; both outcomes are informative.

This paper is a first step rather than a finished pipeline. The pilot IAA study (Section 4) should be extended to a larger sample and could be used as training/evaluation data for ML and LLM-based extraction of RRs. In-context learning approaches to frame-semantic parsing [4] are particularly promising and would directly support large-scale compliance-monitoring tools. In addition, the RRMV mapping (Section 5) should be extended from a single worked example to the full corpus. RRMV's intended successor, the Legal Obligation Metadata Ontology (LOMO)[9], already closes two of the addressed gaps (`hasCondition` on `PrescriptiveRule`, and `isComposedOf` on its result classes), but lacks RRMV's formal versioning, so we treat RRMV Release 1.0.0 as the appropriate target for now.

[8] We stress that this is a single, manually verified proof of concept, not an automated or corpus-wide population. Scaling this to the full corpus is future work (Section 6).

[9] https://drpm.pages.code.europa.eu/lomo/

Data and Code Availability

The annotation guidelines, the annotated corpus (WebAnno TSV3 format), the full Frame/FE-to-RRMV mapping table (JSON and CSV), and the worked RRMV instantiation example (Turtle) described in this paper are available at https://github.com/lecatcel-ops/rr_fn_annotation under a CC BY 4.0 license, consistent with this paper's own licensing.

Acknowledgments

This work builds on the reporting-requests corpus developed within the SORTIS (Study on Regulatory Reporting Standards) project.

Declaration on Generative AI

During the preparation of this work, the author(s) used Claude (Anthropic) to assist with: corpus quality-control analysis of the annotation export; drafting of the annotation guidelines; cross-referencing and mapping frame elements to the RRMV ontology; and drafting of this manuscript's text. The author(s) reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] M. Corazza, M. Palmirani, S. Sapienza, G. Longo, Reporting requests modelling in european legislation with a hybrid ai approach, in: Proceedings of the 20th International Conference on Artificial Intelligence and Law (ICAIL '25), ACM, 2025. URL: https://doi.org/10.1145/3769126.3769255. doi:10.1145/3769126.3769255.

[2] M. Palmirani, D. Liga, Derogations analysis of european legislation through hybrid ai approach, in: A. Kő, E. Francesconi, G. Kotsis, A. M. Tjoa, I. Khalil (Eds.), Electronic Government and the Information Systems Perspective (EGOVIS 2022), volume 13429 of *Lecture Notes in Computer Science*, Springer, Cham, 2022, pp. 123–137.

[3] C. Zheng, X. Chen, R. Xu, B. Chang, A double-graph based framework for frame semantic parsing, in: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Seattle, United States, 2022, pp. 4998–5011.

[4] D. Garat, G. Moncecchi, D. Wonsever, Exploring in-context learning for frame-semantic parsing, 2025. arXiv:2507.23082.

[5] M. Palmirani, R. Sperberg, G. Vergottini, F. Vitali, Akoma ntoso version 1.0. part 1: XML vocabulary, 2018. URL: http://docs.oasis-open.org/legaldocml/akn-core/v1.0/os/part1-vocabulary/akn-core-v1.0-os-part1-vocabulary.html, oASIS Standard, 29 August 2018.

[6] C. F. Baker, C. J. Fillmore, J. B. Lowe, The berkeley framenet project, in: Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL'98), 1998, pp. 86–90.

[7] B. Van Nuffelen, A. Nardin, D. Audrito, C. Guasch, M. Palmirani, Reporting requirement metadata vocabulary (RRMV), v0.2.0, 2025. URL: http://data.europa.eu/2qy/rrmv, contributors: A. Potloot, T. Slangen, A. G. Nuzzolese.

[8] M. Palmirani, A. G. Nuzzolese, G. Longo, Reporting requirements ontology for european legislation, in: Electronic Government and the Information Systems Perspective (EGOVIS 2025), volume 16049 of *Lecture Notes in Computer Science*, Springer, 2026, pp. 18–32.

[9] Council of the European Union, Council conclusions of 6 november 2017 on the european legislation identifier (ELI), 2017. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.C_.2017.441.01.0008.01.ENG, 2017/C 441/05, OJ C 441, 22.12.2017, pp. 8–12.

[10] R. Sanderson, P. Ciccarese, B. Young, Web annotation data model, 2017. URL: https://www.w3.org/TR/annotation-model/, w3C Recommendation, 23 February 2017.
[11] C. J. Fillmore, Frame semantics and the nature of language, Annals of the New York Academy of Sciences 280 (1976) 20–32. doi:10.1111/j.1749-6632.1976.tb25467.x.
[12] C. F. Baker, FrameNet: Frame semantic annotation in practice, in: Handbook of Linguistic Annotation, Springer, 2017, pp. 771–811.
[13] J. Ruppenhofer, M. Ellsworth, M. R. L. Petruck, C. R. Johnson, J. Scheffczyk, FrameNet II: Extended Theory and Practice, Technical Report, International Computer Science Institute, 2006.
[14] M. Palmirani, M. Ceci, D. P. Radicioni, A. Mazzei, FrameNet model of the suspension of norms, in: Proceedings of the 13th International Conference on Artificial Intelligence and Law (ICAIL '11), ACM, 2011, pp. 189–193. doi:10.1145/2018358.2018385.
[15] G. Venturi, Semantic annotation of italian legal texts: A FrameNet-based approach, Constructions and Frames 3 (2011) 46–79.
[16] T. Agnoloni, M. Fernández Barrera, M. T. Sagri, D. Tiscornia, G. Venturi, When a FrameNet-style knowledge description meets an ontological characterization of fundamental legal concepts, in: AI Approaches to the Complexity of Legal Systems (AICOL 2009), volume 6237 of *Lecture Notes in Computer Science*, Springer, 2009, pp. 93–112.
[17] J.-C. Klie, M. Bugert, B. Boullosa, R. Eckart de Castilho, I. Gurevych, The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation, in: Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations, Santa Fe, New Mexico, 2018, pp. 5–9.
[18] J. Cohen, A coefficient of agreement for nominal scales, Educational and Psychological Measurement 20 (1960) 37–46.
[19] R. Artstein, M. Poesio, Survey article: Inter-coder agreement for computational linguistics, Computational Linguistics 34 (2008) 555–596.